Abstract Title Page

Title: The Middle School Intervention Project: Use of a Regression Discontinuity Design to Evaluate a Multi-Component Intervention for Struggling Readers in Middle School in Six School Districts

Authors and Affiliations:

Deanne A. Crone, Ph.D. (University of Oregon)
Michael Stoolmiller, Ph.D. (University of Oregon)
Scott K. Baker, Ph.D. (University of Oregon)
Hank Fien, Ph.D. (University of Oregon)

Abstract Body

Background / Context:
 The primary aim of the Middle School Intervention Project (MSIP) is to evaluate the impact of a multi-component intervention for struggling adolescent readers on reading outcomes. The intervention consists of: (1) targeted, Tier 2 reading and (2) school engagement interventions, and (3) data-based-decision-making (DBDM) teams to review and act on student data.
 The rationale for this evaluation is based on three considerations: (a) Although many middle and high school students in the US are not able to meet current achievement expectations (NAEP, 2007), there is a nationwide trend towards increasing graduation requirements. Increasing demands through course and testing requirements without providing students with extensive support to meet these requirements in the middle and high school years, is a formula for increasing dropout rates (Christensen, et al., 2001; Langenfeld et al., 1997); (b) To meet new achievement and graduation objectives and to prevent more students from dropping out of schools, state and local educational agencies (SEAs and LEAs) are implementing reform efforts targeting at-risk students through state and district-wide initiatives; (c) The Oregon Department of Education and five participating districts are engaged in widespread implementation of such a reform effort in 6th -8th grades and are committed to a rigorous evaluation.
 Supporting at-risk students to meet rigorous graduation requirements and reduce school dropout requires solutions that are comprehensive in nature (Christenson et al., 2001; Dynarski et al., 2008; Fasholoa & Slavin, 1998; Torgesen et al., 2007); (1) Academic and adolescent reading interventions must be intense (Ensminger & Slusarick, 1992; Kamil et al., 2008); (2) Because the reasons for poor school engagement can involve emotional, social, and/or academic factors (Finn 1989; 1993), student support plans need to be sufficiently flexible to address multiple dimensions (Christenson, et al., 2001; Finn & Rock, 1997); (3) Virtually every major recommendation for improving reading/engagement outcomes for at-risk students includes data-based decision making as an essential component (Dynarski et al., 2008; Kamil et al., 2008).
 The results of this intervention implementation across the middle grades, accompanied by a rigorous evaluation, will help LEAs and SEAs understand what effects can be achieved under specific implementation circumstances.
 A critical concept in MSIP is that 6th grade students are screened into the targeted, Tier 2 intervention based on 5th grade reading scores. This screening function allows for a regression discontinuity design (RDD). Use of the RDD allows us to evaluate existing practices in schools, rather than randomly assign half of the schools to receive a prescribed intervention, as in a traditional RCT. Thus, this study is also a test of RDD as a feasible means for conducting rigorous evaluation research with SEAs and LEAs.

Purpose / Objective / Research Question / Focus of Study:
 The purpose of this presentation is to demonstrate the use of RDD to evaluate the impact of a multi-component intervention for struggling readers in middle school. We examine the impact of the intervention on the Oregon Assessment of Skills and Knowledge (OAKS), a high-stakes, statewide reading assessment. We will address four research questions.
 1. Is it feasible to implement an RDD with fidelity, in real school settings across multiple schools within multiple school districts? That is, to what extent did schools comply with the requirements of the RDD?
 2. Were 6th grade interventions successful in increasing reading proficiency?

3. Is there significant heterogeneity in intervention effect sizes across the individual schools?

4. Which schools have significant and substantial 6th grade reading intervention effect sizes?

Setting:

This study was conducted in 40 schools within five school districts in Oregon during the 2010-2011 school year. Seventeen schools were middle schools with a 6th-8th grade configuration. Twenty-two schools were K-6 elementary schools, within one district. One school had a PreK-8 grade configuration. The districts range in size from 5,767 to 36,640 students. The elementary and middle schools range in size from 94 to 814 students; in Free and Reduced Lunch rates from 19% to 84%; in English learners from 1% to 48%; and in minority students from 11% to 65%.

Population / Participants / Subjects:

In Year 1, there were a total of 4,945 6th grade students in the study. Students were placed into the intervention or comparison groups at the beginning of the school year. Placement into intervention was based on the student's combined performance on the OAKS reading assessment and a measure of passage reading fluency in fifth grade. Both measures were standardized and averaged to obtain a single cut score for each student. The schools sorted students from least proficient to most, picked a cut point and all students at or below the cut point received an intervention the following fall in their 6th grade school. Schools were able to pick a cut point that allowed them to serve in the intervention those students they believed needed the intervention. However, they were expected to follow guidelines so that between .20 and .80 of students received the intervention to ensure adequate statistical power for the RD design. Schools were allowed to exempt a small proportion, up to 5%, of students, so called wild cards, from the assignment process, for whatever reason, to allow some flexibility in the assignment process. These wild card students were not used in the evaluation of the school's intervention. The percentage of students placed in the intervention group, by school, ranged from 14.3% to 64.7%.

Intervention / Program / Practice:

Because this study is an evaluation of existing practices at each school, each school had latitude in determining how the 3 components of the intervention were configured in their building. This school-level decision making resulted in substantial variability in intervention configuration across schools. However, the three components were expected to align with certain critical features that were common across all of the participating schools. These critical features are listed in Table 1 in Appendix B.

In Year 1, across the 40 schools, reading interventions averaged 27 weeks in length (SD = 6.8) and 4.8 times per week (SD = .3), for an average duration of 214.8 minutes per week (SD = 26.7). The student to teacher ratio in intervention classes was 9.6 (SD = 1.7). Published programs were used in 80% of the intervention classrooms. Of the 23 different published programs used, Language! (26 schools), Rewards (14), Corrective Reading (9), and Read Naturally (9) were used at the most schools. Engagement interventions ran the gamut from tutoring, to check-in/check-out programs to extracurricular activities after school. On average engagement interventions lasted 22.9 weeks (SD = 3.8) and met 2 times per week (SD = 7). The majority of DBDM meetings included an administrator and reviewed multiple sources of data. On average, DBDM teams met for 55 minutes (SD = 14) and discussed 6.5 (SD = 3.3) intervention students per meeting. At many of these meetings, comparison students or students in other grades were also discussed. The purpose of the DBDM meetings was to: (a) collect and summarize ongoing academic achievement data and student engagement data; (b) monitor academic and engagement data and use data to modify instruction or interventions.

Research Design:

We used a clustered regression discontinuity design, with students as the assignment unit, nested within schools, within districts.

Statistical, Measurement, or Econometric Model:

Not applicable. We are not proposing anything new in the analyses of MSIP.

Usefulness / Applicability of Method:

This study demonstrates the utility of using an RDD to conduct an evaluation of existing practices within the limitations imposed by conducting research in real school settings. Research questions 2, 3, and 4 (above) are answered by parameters from multi-level models applied to the MSIP data. Question 2 is answered by the fixed effect estimate of the difference in average reading proficiency between intervention and comparison groups at the RD cut point. Question 3 is answered by the significance and size of the estimated variance of the RD effect across schools. Question 4 is answered by inspecting empirical or fully Bayesian estimates of the individual school level random RD effects and their associated standard errors.

Data Collection and Analysis:

We collected 2 reading proficiency indicators during spring of 5th grade from 6,761 students in 97 elementary schools from 5 districts, an oral reading fluency measure (EasyCBM PRF in 4 districts and DIBELS ORF (6th edition) in 2 districts), and the Oregon Assessment of Knowledge and Skills (OAKS) reading subtest to use for assignment to intervention in the fall of 6th grade.

We used normative information to standardize the OAKS, PRF and ORF. The resulting standardized scores for the reading indicators were averaged to obtain the cut score used for assignment to 6th grade intervention.

By design, each school picked their own cut point. In order to combine more than 1 school for any type of RDD analysis or plot, the cut scores for students within a school were centered around the school's own cut point (Raudenbush & Bryk, 2002, p. 139). Once this was done, all schools had the same cut point, zero, and analyses proceeded.

Findings / Results:

After schools chose cut points and assigned students to intervention or comparison groups, we checked on the proportion of students assigned to intervention, the proportion of students designated as wild cards, the average cut score and the cut point. Overall compliance with the requirements of the RDD design was high. but there was variation at the individual school level. A number of schools had low proportions of students assigned to intervention (less than 20%), and a number of schools had high proportions of wild cards (greater than 5%).

We performed multilevel RD analyses on observed pseudo-covariates to check that assignment was based solely on the cut score. We checked gender, Special Education status, Limited English Proficiency Status and Free or Reduced Lunch status. We found no evidence of RD effects on any of these covariates. We also performed the McCrary test to check for discontinuities in the cut score distribution at the cut point which would suggest deliberate manipulation of intervention assignment status. We found no evidence of a discontinuity.

For our initial analyses with the OAKS, we included only those students who maintained correct assignment status. This amounted to 4,393 students in 40 schools, 95.0% of all students that attended a participating school in the fall.

All pooled analyses using parametric models also included linear and quadratic effects of the cut point as a school level predictor to compensate for the school cut point centering. To capture the RD effects of interest on intercept and slope, we used the typical approach of including a dummy variable indicator for assignment to intervention and an interaction of the dummy

variable with the slope of the cut score. This amounts to separate regressions in the intervention and comparison groups.

We fit 5 linear, multilevel models that were nested and differed in the number and correlations among the random effects but were identical with respect to the specification described previously for the fixed effects. Model 1 was the simplest with 2 uncorrelated random effects, intercept and slope in the comparison group. Model 2 added an uncorrelated random effect for the difference in intervention and comparison group intercepts at the cut point, the RD effect. Model 3 allowed all 3 random effects to be correlated. Model 4 added an uncorrelated random effect for the difference in intervention and comparison group slopes, the RD effect on slope. Model 5 allowed all 4 random effects to be correlated. For all models, we inspected student level residuals to check on the standard assumptions of normality and homogeneity of variance and of the residuals. We also inspected the estimated random effects to check on the standard assumption of a multinormal distribution. None of the residual plots revealed substantial problems with the standard assumptions. However, regardless of the model variations, the mean RD effect across all the schools was positive, non-significant and very close to zero (one tenth of an OAKS point to one twentieth). The OAKS has a population standard deviation in Oregon of 9, so our population averaged RD effect is essentially zero. To guard against undue influence of outliers far from the cut point or nonlinearity, we repeated the models above for students within 2 points on either side of the cut point (93% of total sample). Results were essentially the same.

The variance of the RD effect was significant (model 1 vs. 2; chi-square = 4.60, df =1, p = .0320). The correlations among the 3 random effects also significantly improved model fit (model 2 vs. 3; chi-square = 9.82, df = 3, p = .0202). The variance and correlations for the 4th random effect did not significantly improve the fit of the model (model 3 vs. model 4; chi-square = 1.43, df = 1, p = .2310 and model 4 vs. 5; chi-square = 0.85, df = 3, p = .8374). Thus there is evidence for significant variation in RD effects across schools.

To examine RD effects for individual schools, we computed empirical Bayes (EB) estimates of school level RD effects and their standard errors using standard formulas based on linear multilevel model 5, from the analyses that limited the sample to students within 2 points of the cut point. Eight schools had significant positive effects, and two schools had significant negative effects. The largest absolute residual from regressing the EB RD effects of model 5 with the cut score limited to -2 to 2 on model 5 with the full range of the cut score was .74 raw OAKS points and the correlation between the two sets of scores was .94. Thus, differences in magnitude and direction of effects across the different models were small.

Conclusions:

Results from the first year of the study demonstrate the feasibility of implementing a large-scale RDD evaluation of existing intervention practices in real school settings. Schools used the cut score procedure, without detectable manipulation, to assign students to intervention or comparison. Although, there were exceptions, in large part, schools adhered to the design requirements. At the project level, there was no significant effect of the multi-component intervention on students' scores on the statewide reading assessment. However, there was significant variation in the effect of the intervention across schools. Regardless of the level of significance or direction of effect, all schools had small effects, ranging from about -1.5 to 1.5 raw OAKS points or about -.17 to .17 for standardized effects in model 5 with cut scores limited to -2 to 2. Future analyses will examine the RD effect on other student outcomes, including, grade point average, passage reading fluency, attendance, and a measure of student engagement.

Appendices

Appendix A. References

Christenson, S. L., Sinclair, M. F., Lehr, C. A., & Godber, Y. (2001). Promoting successful school completion: Critical conceptual and methodological guidelines. *School Psychology Quarterly, 16*(4), 468-484.

Dynarski, M., Clarke, L., Cobb, B., Finn, J., Rumberger, R., & Smink, J. (2008). *Dropout prevention: A Practice Guide* (No. NCES 2008-4025). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, US Department of Education.

Ensminger, M. E., & Slusarcick, A. L. (1992). Paths to high school graduation or dropout: A longitudinal study of a first-grade cohort. *Sociology of Education, 65*(2), 95-113.

Fashola, O. S., & Slavin, R. (1998). Effective dropout prevention and college attendance programs for students placed at risk. *Journal of Education for Students Placed at Risk, 3*, 159-183.

Finn, J. (1989). Withdrawing from school. *Review of Educationa Research, 59*, 117-142.

Finn, J. (1993). *School engagement & students at risk*. Buffalo, NY: State University of New York at Buffalo.

Finn, J., & Rock, D. A. (1997). Academic success among students at risk for school failure. *Journal of Applied Psychology, 82*(2), 231-234.

Kamil, M. L., Borman, G., Dole, J., Kral, C. C., Salinger, T., & Torgesen, J. K. (2008). *Improving adolescent literacy: Effective classroom and intervention practices: A Practice Guide* (Report No. NCEE #2008-4027). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, US Department of Education.

Langenfeld, K., Thurlow, M., & Scott, D. (1997). High stakes testing for students: Unanswered

    questions and implications for students with disabilities. Minneapolis, MN: National

    Center on Educational Outcomes. Available at

    http://education.umn.edu/nceo/OnlinePubs/Synthesis26.htm.

National Assessment of Educational Progress. (2007). *The nation's report card: Reading 2007*

    (report No. NCES 2007496). Washington, DC: National Center for Education Statistics,

    Institute Education Sciences, U. S. Department of Education.

Raudenbush, S.W. & Bryk, A.S. (2002). *Hierarchical linear models: Applications and data*

    analysis methods, CA: Sage Publications, Inc.

Torgesen, J. K., Houston, D. D., Rissman, L. M., Decker, S. M., Roberts, G., Vaughn, S., et al.

    (2007). *Academic literacy instruction for adolescents: A guidance document form the*

    Center on Instruction. Portsmouth, NH: RMC Research Corporation, Center on

    Instruction.

Appendix B. Tables and Figures

Table 1.
Critical Features of Each Component of the Intervention

Reading Intervention	School Engagement Intervention	Data-based Decision Teams
1. Procedure for identifying students who need reading support and indentifying specific needs 2. Instructors have received adequate training to deliver intervention 3. Intervention is delivered for equivalent of 20-30 minutes or for at least one full academic term or until student demonstrates sufficient improvement 4. Procedure for monitoring student progress on targeted skills, as defined by each school 5. Reading intervention is distinct and different from what other non-intervention students receive	1. Students check-in with a supportive adult on a regular basis daily, weekly, monthly (depending on the intensity of the child's school engagement needs) 2. Students receive regular, constructive feedback on behavioral and/or academic performance 3. Students are recognized/reinforced for demonstrating improvement and engagement with school 4. Data is collected on student's behavioral and/or psychological engagement in school 5. Intervention can include a menu of options of what is available in the school, and an effort is made to best match the student's need with the intervention options available (e.g. Check-in/Check-out; Social Skills group; Girls' Club; extracurricular sport, etc.	1. Systematic process for reviewing data and making decisions about students at risk for reading or school engagement problems 2. School-based team includes key stakeholders 3. Data team meets at least 1 x / academic term 4. Data reviewed is relevant, formative, and current